```
import pandas as pd
```

```
dataset = pd.read_csv('Mall_Customers.csv')
```

```
dataset.head()
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
dataset.shape
```

```
(200, 5)
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   CustomerID              200 non-null    int64
 1   Gender                  200 non-null    object
 2   Age                     200 non-null    int64
 3   Annual Income (k$)      200 non-null    int64
 4   Spending Score (1-100)  200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```
dataset.describe()
```

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

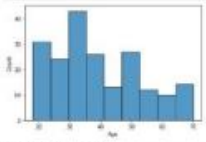
univariant analysis

```
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix,accuracy_score
```

```
import seaborn as sns
```

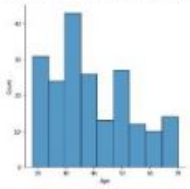
```
import matplotlib.pyplot as plt
```

```
sns.distplot(dataset.Age)
plt.show()
```



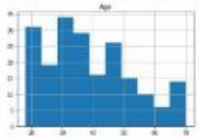

```
sns.distplot(dataset.Age)
```

```
<seaborn.axisgrid.FacetGrid at 0x2b8f9a8b150>
```



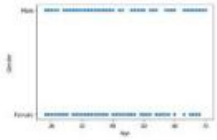

```
dataset.hist('age')
```

```
array([[<AxesSubplot:title={'center':'Age'}>]], dtype=object)
```



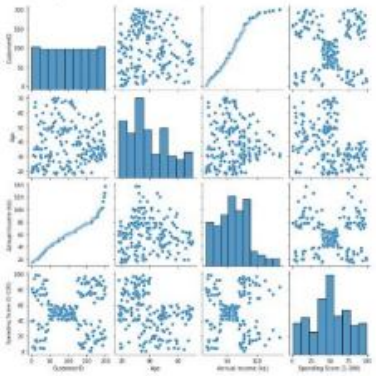

```
sns.scatterplot(x=dataset.Age,y=dataset.Gender)
```

```
<AxesSubplot:xlabel='Age', ylabel='Gender'>
```




```
sns.pairplot(dataset)
```

```
<seaborn.axisgrid.PairGrid at 0x2b8f9f6c0f00>
```



check the missing values and deals withthem

```
dataset.isna()
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
195	False	False	False	False	False
196	False	False	False	False	False
197	False	False	False	False	False
198	False	False	False	False	False
199	False	False	False	False	False

200 rows × 5 columns

```
```